

Intestinal microbiome composition and its relation to joint pain and inflammation

Boer *et al.*

Supplementary Information:

Supplementary Figures:

Supplementary Figure 1: PCA plot of Aitchison distances of inter-individual microbial composition in the Rotterdam Study - Page 2

Supplementary figure 2: Spearman correlation between 16s sequencing and qPCR *Streptococcus* relative abundancies in the Rotterdam Study – Page 3

Supplementary Tables:

Supplementary Table 1: Association results of α -diversity metrics and β -diversity with Knee OA pain and severity in the Rotterdam Study – Page 4

Supplementary Table 2: streptococcus spp. association with WOMAC-knee pain score in Rotterdam Study is independent of alcohol use, smoking and BMI – Page 5

Supplementary Table 3: Streptococcus spp. association with WOMAC-knee pain score in Rotterdam Study. CEU ancestry only- Page 6

Supplementary Table 4: Streptococcus spp. association with WOMAC-knee pain score in Rotterdam Study is not driven by oral medication use – Page 7

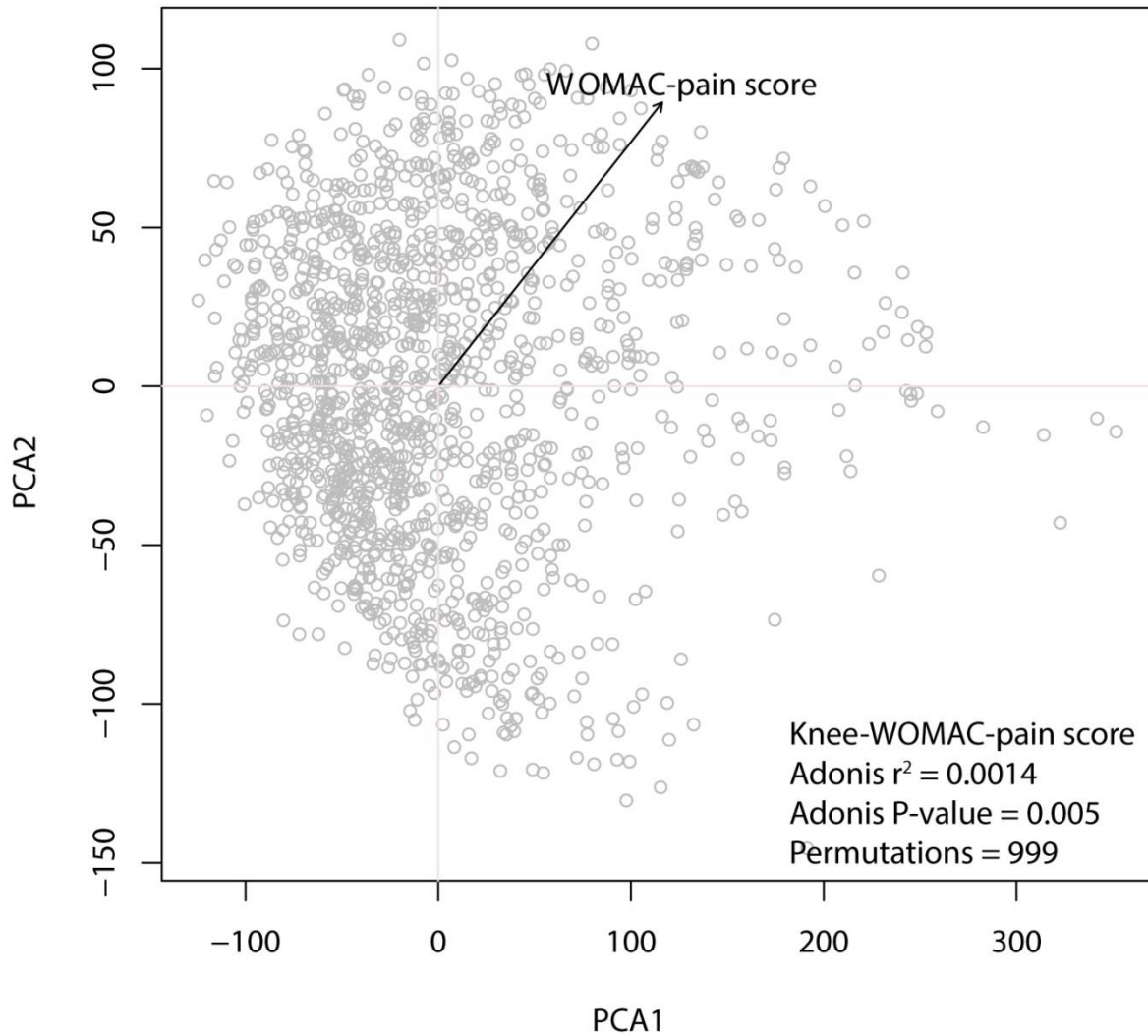
Supplementary Table 5: Increased Streptococcus spp. associated with increased WOMAC-knee-pain- Page 8

Supplementary Table 6: Streptococcus spp. Association with WOMAC-pain score when using ILR and ARS transformation – Page 9

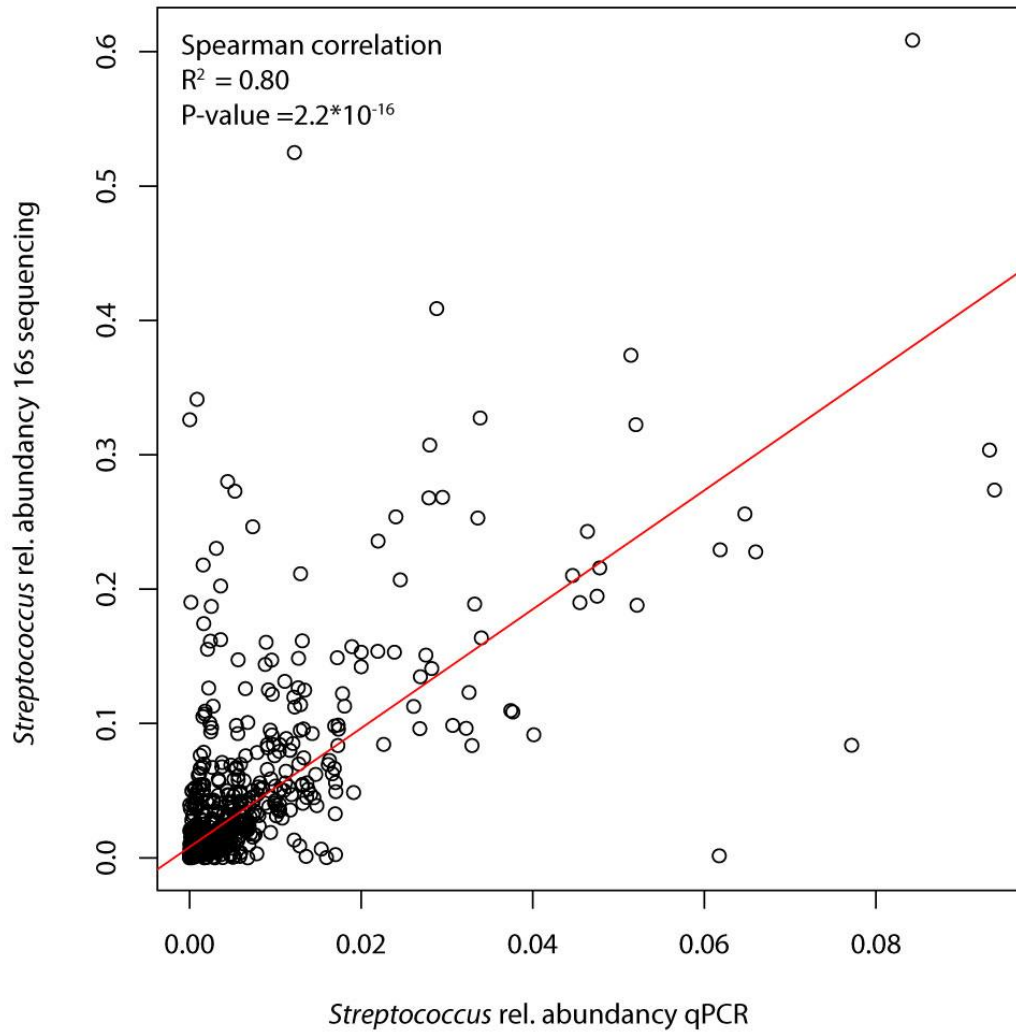
Supplementary Table 7: General characteristics of the replication cohort Lifelines Deep – Page 10

Supplementary Table 8: Increased relative abundancy of Streptococcus is associated with increased Knee WOMAC-pain score, when adjusting for BMI – Page 11

Supplementary Table 9: List of all Primers used in this study – Page 12



Supplementary Figure 1: PCA plot of Aitchison s distances of inter-individual microbial composition in the Rotterdam Study. With plotted vector of knee WOMAC-pain score. PERMANOVA P-value calculated through VEGAN R package ADONIS function, permutations=999, n=1,427



Supplementary Figure 2: Spearman correlation between 16s sequencing and qPCR *Streptococcus* relative abundancies in the Rotterdam Study. Spearman correlation between relative *Streptococcus spp.* abundance determined by 16S rRNA sequencing and quantitative *Streptococcus spp.* abundance determined by qPCR, in the Rotterdam study microbiome. Red line depicts the regression line (16S sequencing abundance ~ qPCR abundance), N=1,427

Supplementary Table 1: Association results of α -diversity metrics and β -diversity with Knee OA pain and severity in the Rotterdam Study

	Knee KLSum		Knee WOMAC pain	
	Beta / r ²	P-Value	Beta / r ²	P-value
Shannon index	-9.5 * 10 ⁻⁰²	0.31	-2.3 * 10 ⁻⁰¹	0.07
Inverse Simpson index	-1.6 * 10 ⁻⁰³	0.69	-8.9 * 10 ⁻⁰³	0.13
B-diversity	1.39 * 10 ⁻⁰³	0.09	1.4 * 10 ⁻⁰³	5.0 * 10 ⁻⁰³

α -diversity metrics used are the Shannon Index, Inverse Simpson Index. Knee KLSum is the sum of the Kellgren-Lawrence score in the left knee and right knee. WOMAC-pain score is a sum of the WOMAC pain questionnaire question for the left and right knee. Poisson regression model was used. For Beta-diversity we calculated the Aitchison distance from CLR normalized data, P-values and r² were calculated using adonis PERMANOVA for 999 iterations. All analysis used were adjusted for adjusted for age, sex, TimeInMail and Batch

Supplementary Table 2: Streptococcus spp. association with WOMAC-knee pain score in Rotterdam Study is independent of alcohol use, smoking and BMI

		Model 1			Model 2			Model 3		
		Age + Sex + Technical CoVar.			Model 1 + Smoking + Alcohol			Model 2 + BMI		
Taxonomy	N	CoE	P-Value	FDR	CoE	P-Value	FDR	CoE	P-Vale	FDR
Class . / <i>Bacilli</i>	1,419	6.07*10 ⁻⁰³	9.06*10 ⁻⁰⁹	2.32*10 ⁻⁰⁶	5.81*10 ⁻⁰³	3.19*10 ⁻⁰⁸	8.16*10 ⁻⁰⁶	5.00*10 ⁻⁰³	2.44*10 ⁻⁰⁶	6.3*10 ⁻⁰⁴
Order .. / <i>Lactobacillales</i>	1,417	6.11*10 ⁻⁰³	7.61*10 ⁻⁰⁹	2.32*10 ⁻⁰⁶	5.85*10 ⁻⁰³	2.65*10 ⁻⁰⁸	8.16*10 ⁻⁰⁶	5.03*10 ⁻⁰³	2.20*10 ⁻⁰⁶	6.3*10 ⁻⁰⁴
Family ... / <i>Streptococcaceae</i>	1,402	4.91*10 ⁻⁰³	1.52*10 ⁻⁰⁷	1.94*10 ⁻⁰⁵	4.70*10 ⁻⁰³	4.52*10 ⁻⁰⁷	5.78*10 ⁻⁰⁵	4.01*10 ⁻⁰³	2.10*10 ⁻⁰⁵	2.7*10 ⁻⁰³
Genus / <i>Streptococcus</i>	1,396	5.04*10 ⁻⁰³	7.28*10 ⁻⁰⁸	1.24*10 ⁻⁰⁵	4.83*10 ⁻⁰³	2.20*10 ⁻⁰⁷	3.76*10 ⁻⁰⁵	4.13*10 ⁻⁰³	1.22*10 ⁻⁰⁵	2.1*10 ⁻⁰³

MaAslin results for Streptococcus WOMAC-pain score association. FDR is based on full dataset (256 taxonomies). Model 1: the base model of WOMAC + age + sex + technical covariates (Technical CoVar: Time in mail + Batch). Model 2 : the base model with Alcohol use and smoking added. Model 3: model 2 with BMI added. T.Covar: technical Covariates, CoE: Coefficient, FDR: Benjamin-Hochberg FDR multiple testing corrected P-value, BMI: Body Mass Index. Analysis done on full dataset (256 taxonomies). P-value, FDR and CoE from MaAslin analysis.

Supplementary Table 3: Streptococcus spp. association with WOMAC-knee pain score in Rotterdam Study. CEU ancestry only.

		Model 1			Model 2			Model 3		
		Age + Sex + Technical CoVar.			Model 1 + Smoking + Alcohol			Model 2 + BMI		
Taxonomy	N	CoE	P-Value	FDR	CoE	P-Value	FDR	CoE	P-Vale	FDR
Class . / <i>Bacilli</i>	1,257	5.83*10 ⁻⁰³	7.58*10 ⁻⁰⁸	9.70*10 ⁻⁰⁶	5.67*10 ⁻⁰³	1.54*10 ⁻⁰⁷	1.98*10 ⁻⁰⁵	5.11*10 ⁻⁰³	3.06*10 ⁻⁰⁶	3.92*10 ⁻⁰⁴
Order .. / <i>Lactobacillales</i>	1,255	5.87*10 ⁻⁰³	6.46*10 ⁻⁰⁸	9.70*10 ⁻⁰⁶	5.72*10 ⁻⁰³	1.30*10 ⁻⁰⁷	1.98*10 ⁻⁰⁵	5.14*10 ⁻⁰³	2.77*10 ⁻⁰⁶	3.92*10 ⁻⁰⁴
Family ... / <i>Streptococcaceae</i>	1,243	5.06*10 ⁻⁰³	2.98*10 ⁻⁰⁷	2.54*10 ⁻⁰⁵	4.94*10 ⁻⁰³	5.65*10 ⁻⁰⁷	4.82*10 ⁻⁰⁵	4.43*10 ⁻⁰³	9.39*10 ⁻⁰⁶	8.02*10 ⁻⁰⁴
Genus / <i>Streptococcus</i>	1,237	4.99*10 ⁻⁰³	9.05*10 ⁻⁰⁷	5.79*10 ⁻⁰⁵	4.86*10 ⁻⁰³	1.66*10 ⁻⁰⁶	1.06*10 ⁻⁰⁴	4.30*10 ⁻⁰³	2.91*10 ⁻⁰⁵	1.86*10 ⁻⁰³

MaAslin results for Streptococcus WOMAC-pain score association on individuals with CEU ancestry only. CEU (central european ancestry) was determined by genetic principal components. Individuals with a minimum of 50% CEU ancestry were included. FDR is based on full dataset (256 taxonomies). Model 1: the base model of WOMAC + age + sex + technical covariates (Technical CoVar: Time in mail + Batch). Model 2 : the base model with Alcohol use and smoking added. Model 3: model 2 with BMI added. T.Covar: technical Covariates, CoE: Coefficient, FDR: Benjamin-Hochberg FDR multiple testing corrected P-value, based on full dataset (256 taxonomies), BMI: Body Mass Index. P-value, FDR, CoE from MaAslin analysis.

Supplementary Table 4: Streptococcus spp. association with WOMAC-knee pain score in Rotterdam Study is not driven by oral medication use

Taxonomy	N	Model 3		Model 3		Model 3		Model 3		
		+ NSAID use		+ PPI use		+ NSAID use + PPI use		+ Exclusion of PPI users		
		CoE	P-value	CoE	P-value	CoE	P-value	N	CoE	P-value
Class . / <i>Bacilli</i>	1,419	5.09*10 ⁻⁰³	2.04*10 ⁻⁰⁶	4.20*10 ⁻⁰³	4.06*10 ⁻⁰⁵	4.16*10 ⁻⁰³	5.25*10 ⁻⁰⁵	1,123	3.25*10 ⁻⁰³	7.96*10 ⁻⁰⁴
Order .. / <i>Lactobacillales</i>	1,417	5.11*10 ⁻⁰³	1.90*10 ⁻⁰⁶	4.22*10 ⁻⁰³	3.76*10 ⁻⁰⁵	4.18*10 ⁻⁰³	4.89*10 ⁻⁰⁵	1,121	3.31*10 ⁻⁰³	6.29*10 ⁻⁰⁴
Family ... / <i>Streptococcaceae</i>	1,402	4.06*10 ⁻⁰³	1.98*10 ⁻⁰⁵	3.32*10 ⁻⁰³	2.65*10 ⁻⁰⁴	3.27*10 ⁻⁰³	3.61*10 ⁻⁰⁴	1,109	2.38*10 ⁻⁰³	3.67*10 ⁻⁰³
Genus / <i>Streptococcus</i>	1,396	4.17*10 ⁻⁰³	1.22*10 ⁻⁰⁵	3.43*10 ⁻⁰³	1.66*10 ⁻⁰⁴	3.37*10 ⁻⁰³	2.38*10 ⁻⁰⁴	1,104	2.40*10 ⁻⁰³	3.97*10 ⁻⁰³

MaAslin results for Streptococcus WOMAC-pain score association. Model 3: base model of WOMAC + age + sex + technical covariates with smoking, alcohol consumption and BMI added. T.Covar: technical Covariates, CoE: Coefficient, BMI: Body Mass Index. NSAID: non-steroidal anti-inflammatory drugs, and PPI: Proton Pump Inhibitors. N: nr of individuals in the analysis. P-value and CoE from MaAslin analysis.

Supplementary Table 5: Increased *Streptococcus* spp. associated with increased WOMAC-knee-pain score is not driven by WOMAC distribution or outliers

Taxonomy	WOMAC-Pain (>0)					WOMAC-Pain (<10)				
	N	Model 1		Model 3		N	Model 1		Model 3	
		CoE	P-value	CoE	P-value		CoE	P-value	CoE	P-value
Class . / <i>Bacilli</i>	313	5.34*10 ⁻⁰³	9.90*10 ⁻⁰³	4.29*10 ⁻⁰³	3.97*10 ⁻⁰²	1,392	5.67*10 ⁻⁰³	2.00*10 ⁻⁰⁵	4.49*10 ⁻⁰³	7.40*10 ⁻⁰⁴
Order .. / <i>Lactobacillales</i>	312	5.34*10 ⁻⁰³	1.01*10 ⁻⁰²	4.27*10 ⁻⁰³	4.10*10 ⁻⁰²	1,390	5.70*10 ⁻⁰³	1.88*10 ⁻⁰⁵	4.50*10 ⁻⁰³	7.29*10 ⁻⁰⁴
Family ... / <i>Streptococcaceae</i>	309	6.06*10 ⁻⁰³	7.51*10 ⁻⁰⁴	5.40*10 ⁻⁰³	3.14*10 ⁻⁰³	1,375	4.39*10 ⁻⁰³	3.13*10 ⁻⁰⁴	3.41*10 ⁻⁰³	5.31*10 ⁻⁰³
Genus / <i>Streptococcus</i>	306	6.25*10 ⁻⁰³	5.29*10 ⁻⁰⁴	5.56*10 ⁻⁰³	2.38*10 ⁻⁰³	1,369	4.53*10 ⁻⁰³	2.03*10 ⁻⁰⁴	3.53*10 ⁻⁰³	3.93*10 ⁻⁰³

Model 1: WOMAC score knee adjusted for age, sex and technical covariates. Model 3: model 1 adjusted for smoking, alcohol use and BMI. CoE: Coefficient, FDR: Benjamin-Hochberg FDR multiple testing corrected P-value, based on full analysis (248 taxonomies) in MaAslin, N = nr of individuals included in the analysis

Supplementary table 6: Streptococcus spp. Association with WOMAC-pain score when using ILR and ARS transformation

			Model 1		Model 2		Model 3	
			Age + Sex + Technical CoVar		Model 1 + Alcohol + Smoking + BMI		Model 3 + NSAIDs + PPI	
Taxonomy	Method	N	CoE	P-value	CoE	P-value	CoE	P-value
Genus / <i>Streptococcus</i>	ARS	1,395	5.04*10 ⁻⁰³	7.28*10 ⁻⁰⁸	4.13*10 ⁻⁰³	1.22*10 ⁻⁰⁵	2.40*10 ⁻⁰³	3.97*10 ⁻⁰³
Genus / <i>Streptococcus</i>	ILR	1,395	9.28*10 ⁻⁰²	9.87*10 ⁻⁰⁶	7.01*10 ⁻⁰²	8.35*10 ⁻⁰⁴	5.38*10 ⁻⁰²	8.49*10 ⁻⁰³

ARS: arcsine square root transformation applied in MaAslin analysis, ILR: isometric log ratio transformation, ILR was done on full genus level taxonomy dataset. CoE: coefficient, Technical CoVar: technical covariates consisting of TimeInMail and Batch. P-value and CoE, from MaAslin analysis.

Supplementary Table 7: General characteristics of the replication cohort Lifelines Deep

General Characteristics	Females	Males	Total
Cohort participants	505	362	867
Age (years)	45.6(12.9)	45.7(13.2)	45.6(13)
BMI (Kg/m²)	25.1(4.6)	25.4(3.4)	25.2(4.2)
Alcohol (gram/day)	6.2(6.6)	10.8(11.1)	8.1(9)
Smoking (y/n)	87 smokers/416 non-smokers	58 smokers/303 non-smokers	145 current smokers
PPI (y/n)	42 users/444 non-users	22 users/326 non-users	64 current PPI users
NSAIDs (y/n)	17 users/469 non-users	11 users/337 non-users	28 current NSAID users
Knee Phenotypes			
Knee OA (y/n)	x	x	x
KLSum score	x	x	x
WOMAC-Pain score	1.2(2.8)	0.5(1.7)	0.9(2.4)
WOMAC-Pain score > 0	133	64	197
α-diversity metrics			
Shannon Index	3.8(0.4)	3.7(0.3)	3.7(0.4)
Inverse Simpson Index	22.2(8)	20.9(7.2)	21.7(7.7)

PPI: Proton Pump Inhibitor, BMI: Body Mass Index, NSAID: non-steroidal anti-inflammatory drugs, Klscore: Kellgren-Lawrence sum score of the knee joints

Supplementary Table 8. Increased relative abundance of Streptococcus is associated with increased Knee WOMAC-pain score, when adjusting for BMI

Taxonomy	%	N	Rotterdam Study			Life Lines Deep			Meta-Analysis		
			CoE	SE	P-Value	N	CoE	SE	P-value	N	P-Vale
Class . / <i>Bacilli</i>	27.3%	1,419	5.17*10 ⁻⁰³	1.06*10 ⁻⁰³	1.21*10 ⁻⁰⁶	867	5.14*10 ⁻⁰³	1.30*10 ⁻⁰³	8.47*10 ⁻⁰⁵	2,286	3.5*10 ⁻¹⁰
Order .. / <i>Lactobacillales</i>	100%	1,417	5.20*10 ⁻⁰³	1.06*10 ⁻⁰³	1.10*10 ⁻⁰⁶	864	4.53*10 ⁻⁰³	1.33*10 ⁻⁰³	6.82*10 ⁻⁰⁴	2,281	2.6*10 ⁻⁰⁹
Family ... / <i>Streptococcaceae</i>	79.6%	1,402	4.16*10 ⁻⁰³	9.43*10 ⁻⁰⁴	1.12*10 ⁻⁰⁵	863	2.72*10 ⁻⁰³	1.28*10 ⁻⁰³	3.47*10 ⁻⁰²	2,265	1.5*10 ⁻⁰⁶
Genus / <i>Streptococcus</i>	98.7%	1,396	4.28*10 ⁻⁰³	9.45*10 ⁻⁰⁴	6.41*10 ⁻⁰⁶	860	3.03*10 ⁻⁰³	1.59*10 ⁻⁰³	5.74*10 ⁻⁰²	2,256	1.1*10 ⁻⁰⁶

MaAslin analysis results, adjusted for age, sex, technical covariates: DNA isolation batch and time In mail, and BMI. Rotterdam Study (n=1,427), Life Lines Deep (n=867), Meta-analysis: Rotterdam Study and Life Lines Deep together, weighted meta-analysis in METAL. Taxonomy% = percentage of taxonomy is from one taxonomy level higher, ex. 23.7% of all Firmicutes are Bacilli. N = number of individuals in cohort where microbial abundance is not zero for that taxonomy. FDR: P-value adjusted for multiple testing, Benjamin-Hochberg False Discovery Rate, CoE: Coefficient, SE: Standard Error.

Supplementary Table 9: List of all primers used in this study

Primer	Sequence	Target	Ref
319F	ACTCCTACGGGAGGCAGCAG	V3 region	¹
515F	GTGCCAGCMGCCGCGGTAA	V4 region	²
806R	GGACTACHVGGGTWTCTAAT	V4 region	^{1,2}
U16SRTF	ACTCCTACGGGAGGCAGCAGT	16S	³
U16SRTR	TATTACCGCGGCTGCTGGC	16S	³

V3/V4 regions: hypervariable region 3,4 of the bacterial 16S rRNA gene, 16S: bacterial 16S rRNA gene.

References:

1. Fadrosch, D. W. et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2, 6 (2014).
2. Fu, J. et al. The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. *Circ. Res.* 117, 817–824 (2015).
3. Clifford, R. J. et al. Detection of Bacterial 16S rRNA and Identification of Four Clinically Important Bacteria by Real-Time PCR. *PLoS One* 7, e48558 (2012).